

Remarks

Claims 1-11 remain pending in this application after entry of this paper. Applicant has amended the title to more specifically indicate the invention to which the claims are directed.

Claim 1 is an independent claim and recites a method for converting text to concatenated voice by utilizing a digital voice library and a set of playback rules. The digital voice library includes a plurality of speech items including words and syllables. The digital voice library further includes a corresponding plurality of voice recordings. Each speech item corresponds to at least one available voice recording. The method comprises training the digital voice library to associate each syllable speech item with a literal text syllable of the particular syllable speech item.

This is exemplified in Figures 6-7. The prior art fails to suggest this specifically recited combination including the association of each syllable speech item with a literal text syllable of the particular syllable speech item.

Cecys does describe utilization of multiple voice sources in a speech synthesizer. Cecys describes a speech synthesizer with the capability to select among and between a multiplicity of voice sources to provide a higher quality and greater variety of possible synthetic speech sounds. Cecys fails to describe or suggest the association of each syllable speech item with a literal text syllable.

Cecys does describe making a mapping between the phonemes to be spoken and the duration of the subdivisible segments of the recorded sound samples to be used as voice sources for each phonemes. Cecys also describes combining phonemes into syllables and then operating on a syllable by syllable basis, rather than operating on a phonemes by phonemes basis.

Nevertheless, Cecys offers no suggestion of the specifically recited combination in claim 1 involving the training of the digital voice library and the association of each syllable speech item with a literal text syllable of the particular syllable speech item. Cecys is not about training a digital voice library in the particular claimed way. Cecys is about utilization of multiple sources in the speech synthesizer and fails to suggest the claimed invention.

The Examiner makes reference to column 11, lines 4-25 of Cecys. This portion of Cecys only describes combining phonemes into syllables and operating on a syllable by syllable basis to construct the synthesized speech. Cecys fails to suggest the claimed invention and only mentions that a word has a phonetic equivalent which may be processed according to the Cecys technique. This offers nothing about training a digital voice library to associate each syllable speech item with a literal text syllable of the particular syllable speech item.

For the reasons given above, claim 1 is believed to be patentable.

Claim 2 is believed to be separately patentable from claim 1. Claim 2 recites receiving a sequence of words including known words that correspond to word speech items in the digital voice library and including unknown words. Each known word is converted into a word speech item in accordance with the digital voice library. For an unknown word, the unknown word is parsed to determine a sequence of literal text syllables. The text syllable sequence is converted to a sequence of syllable speech items in accordance with the digital voice library. Claim 2 recites an innovative technique for handling unknown words in a method for converting text to concatenated voice. The parsing of an unknown word to determine the sequence of literal text syllables, and the converting of the text syllable sequence to a sequence of syllable speech items in accordance with the digital voice library, in the recited combination, are not suggested by the prior art.

The Examiner has relied on Parthasarathy as a secondary reference in rejecting claims 2-4. Parthasarathy only describes identifying a speaker using mixture discriminant


analysis to develop speaker models. This has nothing to do with the claimed invention. After all, Parthasarathy is about uttered passwords, and involves the constructing of speaker models.

The Examiner makes reference to enrollment and training phases. These features fail to overcome the shortcoming of the primary reference, make no suggestion of associating each syllable speech item with the literal text syllable of the particular speech item as recited by claim 1, and also do not suggest the parsing and converting recited by claim 2.

After all, enrollment involves storing a phone string of a speaker's password utterance, and the training phase only describes developing models.

The remaining claims are dependent claims and are also believed to be patentable.

Respectfully submitted,
ELIOT M. CASE

By 
Jeremy J. Curcuri
Reg. No. 42,454
Attorney for Applicant

Date: June 2, 2005

BROOKS KUSHMAN P.C.
1000 Town Center, 22nd Floor
Southfield, MI 48075-1238
Phone: 248-358-4400
Fax: 248-358-3351